



BEFORE THE BOARD OF APPEALS AND INTERFERENCES  
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Serial No. 09/865,090

Customer No. 23379

Applicant: Harold R. Garner

Confirmation No: 2902

Filed: May 24, 2001

Group Art Unit: 2177

Docket No. UTSD: 0668

Examiner: Wong, Leslie

Title: *A Program for Microarray Design and Analysis*

CERTIFICATE OF MAILING

I hereby certify that this corr. is being deposited with the US Postal Service as First Class Mail in an envelope addressed to the Comm. for Patents, PO Box 1450, Alexandria, VA 22313-1450 on January 3, 2005.

Signed

  
Richard Aron Osman

BRIEF ON APPEAL

The Honorable Board of Appeals and Interferences  
United States Patent and Trademark Office  
P.O. Box 1450  
Alexandria, VA 22313-1450

Dear Honorable Board:

We appeal from the Examiner's Aug 3, 2004 final rejection of claims 1-24.

REAL PARTY IN INTEREST

The real party in interest is The University of Texas System Board of Regents, the assignee of this application.

RELATED APPEALS AND INTERFERENCES

Appellants are unaware of any related appeals or interferences.

STATUS OF THE CLAIMS

Claims 1-24 are rejected and subject to this appeal.

STATUS OF THE AMENDMENTS

All Amendments are believed to be properly before the Board.

## SUMMARY OF THE INVENTION

The invention relates to computer-based systems and corresponding methods for the design and analysis of biopolymer sequence arrays. (Specification, p.5, lines 22-23)

In a first principal embodiment, the invention provides a computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding annotations, the system comprising:

a) a search function which searches the annotations of the dataset according to a user-defined criterion and outputs a first subset of the dataset restricted by the criterion;

b) a redundancy reducing function which compares the first subset with a first database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputs a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;

c) a selection function which applies to the second subset a user-defined selection parameter and outputs a third subset restricted relative to the second subset by the parameter; and

d) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset. (Specification, p.5, line 24 - p.6, line 8)

The system may optionally incorporate one or more of the following limitations:

the criterion is selected from the group consisting of a keyword and a concept;

the criterion is one of a plurality of user-defined criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria;

the criterion is one of a plurality of user-defined criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria, wherein the criteria include multiple keywords;

the dataset is selected from the group consisting of GenBank, Medline and KEGG;

the dataset is one of a plurality of datasets, and the search function searches the annotations of the datasets according to the user-defined criterion and outputs a first subset of the datasets restricted by the criterion;

the database is selected from the group consisting of UniGene and LocusLink;

the database is one of a plurality of databases correlating the sequence identifiers of the

first subset with syngeneic biopolymers, and the redundancy reducing function compares the first subset with the databases and outputs the second subset of the dataset;

the parameter is selected from the group consisting of source, species, author and pathway;

the parameter is one of a plurality of user-defined selection parameters, and the selection function applies to the second subset the parameters and outputs the third subset restricted relative to the second subset by the parameters;

the redundancy reducing function outputs a second subset of the dataset which eliminates unique, natural complex biopolymer redundancy relative to the first subset; and

the system further comprises an expansion function which searches a second database for synonyms of the sequence identifiers of the first, second or third subset. (Specification, p.6, lines 8 - p.7, line 2)

In a second principal embodiment, the invention provides a computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the system comprising:

a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset. (Specification, p.7, lines 3-13)

The system may optionally incorporate one or more of the following limitations:

the merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter; and

the merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter, wherein the parameter is selected from the group consisting of source, author and pathway. (Specification, p.7, lines 14-21)

In a third principal embodiment, the invention provides a computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers

corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations. (Specification, p.7, line 22 - p.8, line 1)

The system may optionally incorporate the following limitation:

the second annotations comprise data attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset, said data selected from the group consisting of: gene expression data, sequencing data, genotype data, polymorphism data and clinical data. (Specification, p.8, lines 2-6)

In yet another embodiment, the invention provides a computer-based system incorporating the elements of the first, second, and optionally, the third principal embodiments described herein. (Specification, p.8, lines 7-9)

In a particular embodiment, the recited systems and methods have been implemented in a computer tool called ARROGANT. This program has been developed to facilitate the identification, analysis and comparison of collections of genes or clones. ARROGANT, in the analysis mode, is a comprehensive tool for providing annotation to large gene collections. ARROGANT takes in a large collection of gene identifiers and associates it with other information collected from many sources like sequence annotations, pathways, homology, polymorphisms, artifacts etc. to help the researcher draw scientific conclusions, understanding, and proceed with future experiments. The simultaneous annotation for a large assembly of genes makes the collection of genomic / EST sequences truly informative. For example, if the collection of genes is used for microarrays, ARROGANT predicts cross-hybridization with the members on the array and the entire UniGene database to help the researcher to design probes that avoid cross-hybridization or alerts the user of their presence. In the design mode, ARROGANT assists in compiling a gene collection, using several different databases simultaneously, queried with keywords and their synonyms. ARROGANT, in one integrated package, also facilitates the design of expression / resequencing microarrays by designing

primers, looking for commercially available clones and designing probes for resequencing. The package also has a third mode of operation to eliminate sequence redundancies and duplicates from multiple gene collections. This is very useful in identifying redundancies due to sequences or clones having different accession numbers but representing fragments of the same gene. This simplifies comparing experiments from various research groups. ARROGANT has been successfully applied to many large gene collections for microarrays, complex multigenic trait projects, polymorphism discovery projects etc. (Specification, p.8, line 10 - p.9, line 1)

### ISSUES

- I. WHETHER THE EXAMINER HAS PROPERLY REJECTED CLAIMS 1, 3, 6 and 8-24 UNDER 35USC103(a).
- II. WHETHER THE EXAMINER HAS PROPERLY REJECTED CLAIMS 2 AND 4 UNDER 35USC103(a).
- III. WHETHER THE EXAMINER HAS PROPERLY REJECTED CLAIM 5, UNDER 35USC103(a).
- IV. WHETHER THE EXAMINER HAS PROPERLY REJECTED CLAIM 7, UNDER 35USC103(a).

### GROUPING OF THE CLAIMS

For Issue I, claims 1, 3, 6, 7, 13 shall stand as a group; and each of claims 8-12 and 14-24 shall stand individually.

For Issue II, each of claims 2 and 4 shall stand individually.

For Issue III, claim 5 shall stand as a group.

For Issue IV, claim 7 shall stand as a group.

### ARGUMENT

- I. THE EXAMINER HAS NOT PROPERLY REJECTED CLAIMS 1, 3, 6 and 8-24 UNDER 35USC103(a).

In her first Action of Oct 31, 2003, the Examiner rejected our claims over Ford in view of

Chin and MacLeod. With no amendments to our claims, the Examiner withdrew those rejections and now cites Wolffe (US 2002/0081603 A1) in view of Hennig (A data-analysis pipeline for large-scale gene expression analysis, 2000, AMC pages 165-173), Lincoln (US Patent 6,303,297), Chin (US Patent 6,470,277) and MacLeod (US Patent 6,221,600 B1). We appreciate that the claimed subject matter is arcane and not easy to examine; however, we believe that the newly cited art similarly does not provide a remotely colorable suggestion of the subject claims.

All our claims recite a highly-specialized computational system for creating a targeted collection of sequences from a dataset or plurality of datasets of sequence identifiers corresponding to natural complex biopolymer sequences which may be linked to corresponding annotations. For instance, in a practical application, the targeted collection of sequences may be used to assemble cDNA sequences for a particular gene expression microarray. Accordingly, the system of our representative claim 1 must provide all four of the following functions:

- a) a search function which searches the annotations of the dataset according to a user-defined criterion and outputs a first subset of the dataset restricted by the criterion;
- b) a redundancy reducing function which compares the first subset with a first database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputs a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;
- c) a selection function which applies to the second subset a user-defined selection parameter and outputs a third subset restricted relative to the second subset by the parameter; and
- d) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset.

The cited art does not teach or suggest the claimed computational tool. We will get into detail below, but in essence, Wolffe et al. (US 2002/0081603 A1) describe methods for characterizing DNA sequences, and disclose that known computer-based methods, such as alignment tools, can be used to compare identified regions with known sequences. Aside from the specific deficiencies noted by the Examiner, the Wolffe protocol is neither applicable nor germane to the field of our invention (and conversely, our invention is neither applicable nor particularly germane to his). Wolffe identifies accessible genomic sequences, and characterizes them as regulatory sequences using known alignment algorithms. We disclose and claim a novel protocol for generating targeted collections (e.g. sequence arrays) of sequences from a dataset of

sequence identifiers corresponding to natural complex biopolymer sequences (e.g. syngeneic sequences) and linked to corresponding annotations.

No amount of supplementing or modifying is going to transform Wolffe's methods into our invention. Of course (as observed by the Action, p.3, line 17 - p.4, line 2), Wolffe does not provide for reducing redundancy of initial search results by mapping to a database correlating sequence identifiers with syngeneic biopolymers to generate a second dataset subset (our claim 1, step (b)). Wolffe is characterizing sequences – Wolffe is not in the business of generating syngeneic datasets. Hence, Wolffe necessarily has no provision for further processing a resultant second dataset subset, as required by our claim 1, steps (c) - (d). Note that analogous required steps for reducing redundancy of the initial search results by mapping to a database correlating sequence identifiers with syngeneic biopolymers are present in all of our claims (e.g. step (b) of claim 13, and step (a) of claims 14, 17, 18 and 20).

How is it possible to transform a method of characterizing regulatory sequences using alignment tools into the claimed method of generating targeted collections (e.g. sequence arrays) of sequences from a dataset of sequence identifiers corresponding to natural complex biopolymer sequences (e.g. syngeneic sequences) and linked to corresponding annotations?

Hennig et al. (2000, Annual Conference on Research in Computational Molecular Biology p.165-173; *INVITED PRESENTATION: A data-analysis pipeline for larger-scale gene expression analysis*) describe a method for characterizing cDNA clone libraries based on oligo fingerprints (OFPs). In this method, EST clones are amplified by PCR, immobilized on filter membranes, and hybridized in separate, parallel incubations to different, known-sequence radiolabeled oligo probes, providing corresponding different hybridization signals for each clone – an oligo fingerprint. Hennig, p.166, first full para.

Oligo fingerprints can be used to identify a subset of low redundant EST clones for genome sequencing efforts: specialized algorithms can be used to cluster clones according to oligo fingerprints and then representative clones from each cluster can be selected to generate a less redundant EST set, which will (hopefully) be representative of the original EST libraries in terms of containing representatives of all the originally represented genes. In theory, such a subset reduces the number of clones which need to be sequenced (Hennig, p.166, second full para), though in practice, the method is quite imperfect (Hennig, p.170, first full para.).

How does the practitioner of Wolffe find applicable relevance in Hennig, and to what end? Wolffe is characterizing novel regulatory sequences by using alignment tools to compare

them with known sequences. Hennig is characterizing large EST libraries based on oligo fingerprinting, so as to reduce the number of clones that need to be sequenced. The Action proposes that Hennig's teachings would have allowed Wolffe to clean, remove duplicates, and perform quality checks to the raw sequence in preparation for the sequence comparison analysis. Action, p.4, lines 10-12. Clean what? Remove duplicates of what? Perform quality checks on what raw sequence? The proposed combination does not tolerate any scrutiny.

Wolffe compares his identified sequences with reference sequences such as in Genbank to generate "hits", such as by using the BLAST algorithm. Of course, to the extent a Wolffe practitioner is generating original sequence, she may well seek to improve the relevance of her sequencing by sequencing multiple sample copies, and using algorithms to identify and discount artifactual sequences. This is not really analogous to what Hennig is doing: spotting duplicate probes to insure accuracy of each probe-EST correlation. But it could be argued to be general motivation to repeat data points and improve accuracy. However, as much as coopting Hennig's data cleaning, removing duplicates, and performing quality checks may improve accuracy, it has not driven Wolffe's practitioner into a different line of work.

To underscore this analysis, we further dissect the cited art, particularly the portions specifically cited in the Action. Wolffe describes methods for identifying, isolating and characterizing regulatory DNA sequences (Abstract, first line). Beginning in Section 0340, Wolffe teaches that computer-based methods can be used to compare identified regions with other sequences, such as known regulatory regions; Wolffe, p.31, para 5. The Action specifically cites sections 0350, 0358, 0386, 0391, 0392, 0397 and 0398; Action, p.3, lines 8-16.

Section 0350 teaches that sequence comparisons can be conducted using known sequence comparison algorithms; Wolffe, p.32, para 5.

Section 0358 teaches that the computer system can implement the comparison by retrieving sequence from an internal database, comparing such sequence with reference sequence using alignment algorithms, and displaying the results for user viewing; Wolffe, p.33, para 4.

Section 0386 describes the "sequence" table 152 of Figure 21, which includes identified sequences to be compared with other sequences, such as known regulatory sequences. Each sequence is represented by a distinct identifier, and can be associated with additional attributes, such as sequence length, BLAST values, etc; Wolffe, p.36, para 1.

Section 0391 describes the "project" table 154 of Figure 21, which includes attributes for sequences identified as being common or unique one or more libraries. This database can also



include fields for describing the sequences by prospective function as predicted by sequence comparison, such as potential positive regulatory sequences, potential negative regulatory sequences, etc; Wolffe, p.36, para 6.

Section 0392 describes an optional “external hit” table 156, which summarizes hits (matches) against sequences stored in public sequence databases, such as Genbank. Typically, each record in this table includes a hit ID and a hit description to identify the matched sequence. Analogously, the database can include an “internal hit” table 158 to summarize hits against sequences of an internal database; Wolffe, p.36, para 7.

Sections 0395 - 0399 describe the graphical user interface of the computer systems. In particular, section 0397 describes a project information button to view a screen to input a project identifier as a query. The computer then retrieves information about the selected project, such as sequence information, the source of the original chromatin sample, and hits against internal and external databases. This project information screen can also be used to input query data or parameters, such as a clone identifier, and retrieve a list of projects that include such data or parameters; Wolffe, p.37, para 1.

The final cited section 0398 describes a sequence database button which allows a user to input sequence identifiers to retrieve polynucleotide sequence information. This button also provides screens to conduct various sequence alignments, such as BLAST, against sequences of internal or external databases, and screens to view alignments; Wolffe, p.37, para 2.

In sum, the cited sections of Wolffe teach that well-known computer-based methods, such as BLAST, can be used to compare identified gene sequences with known sequences, such as known regulatory sequences. So exactly what in Wolffe pertains to the claimed methods of creating sequence collections or arrays from correlated datasets?

Reconsider our representative claim 1: A computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding annotations, the system comprising:

- a) a search function which searches the annotations of the dataset according to a user-defined criterion and outputs a first subset of the dataset restricted by the criterion;

- b) a redundancy reducing function which compares the first subset with a first database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputs a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;

c) a selection function which applies to the second subset a user-defined selection parameter and outputs a third subset restricted relative to the second subset by the parameter; and  
d) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset.

Note that all elements of the claim are inter-related, wherein the output of one step is specifically recited as the input of the next. Hence, the claimed system is limited to a sequential series of elements that make sense only in relation to the antecedent elements. Note that even the search function of element (a) recites and relates back to the dataset of the preamble; hence, even the first recited element is defined by and requires the context of the antecedent preamble, and vice versa: “the” dataset of element (a) is the source of the created targeted collection of sequences of the preamble.

Each subsequently recited element similarly requires antecedents in prior elements. For example, the first element (a) must output a first subset of preamble-recited dataset restricted by the criterion of the search function, and it is this element (a)-recited first subset output which must be subject to the redundancy reducing function of element (b). Our claim must be read as a whole, as a system of specifically interrelated elements. We do not claim *any* combination of database search, redundancy reducing, selection and tabulation functions. Note that the Action-created method of steps (a), (c) & (d) lacks antecedents and does not make any logical sense; for example, from where comes the second subset recited in step (c)?

The cited art does not support any reasoned comparison to our claimed invention. Hence, the Action merely asserts “Wolffe teach” and then copies word-for-word from our claim 1 (Action, p.3, line 4-16). No citation is attempted for our preamble which limits the claimed subject matter and recites the original antecedents of subsequent elements. Furthermore and tellingly, there is no numerical order to the subsequent cited sections of Wolffe (0397, 0398; 0358, 0386; and 0350, 0391 and 0392); in fact, the subsequent citation are asymmetrically extracted out of Wolffe’s order according to the template of our claim. Even so contrived, a careful review (*supra*) of these cited Wolffe sections (in order: 0350, 0358, 0386, 0391, 0392, 0397 and 0398) confirms that Wolffe is comparing identified sequences with known, database sequences by alignment. Nowhere does the Action point to a protocol described by Wolffe that is in any way comparable to the recited interrelated elements (a), (c) & (d) of our claim 1.

Like the previously cited Ford et al. (US Pat No. 6,472,173), the computer-based systems

described by Wolffe are conventional search and alignment protocols, such as found in BLAST. As observed also by the Action (p.3, line 17 - p.5, line 2), such protocols do not go past the initial search function (our claim 1, step (a)) of our method; for example, there is no provision in BLAST, etc., for reducing redundancy of the search results by mapping to a database correlating sequence identifiers with syngeneic biopolymers to generate a second dataset subset (our claim 1, step (b)); and hence, no provision for further processing the resultant second dataset subset, as required by our claim 1, steps (c) - (d). Note that analogous required steps for reducing redundancy of the initial search results by mapping to a database correlating sequence identifiers with syngeneic biopolymers are present in all of our claims (e.g. step (b) of claim 13, and step (a) of claims 14, 17, 18 and 20).

No amount of supplemental citations is going to suggest swapping Wolffe's disclosure for that of ours. As discussed above, Hennig describe a method for characterizing cDNA clone libraries based on oligo fingerprints (OFPs). In this method, EST clones are amplified by PCR, immobilized on filter membranes (~25,000 different clones per filter), and hybridized in separate, parallel incubations to 200-300 different, known-sequence radiolabeled oligo probes, providing corresponding 200-300 different hybridization signals for each clone – an oligo fingerprint (Hennig, p.166, first full para). Oligo fingerprints can theoretically be used to help generate a unique set of sequences describing the complete gene set of an organism (Hennig, Introduction, lines 1-4). Details of this oligo fingerprint method of EST selection are described in Hennig's subsequent sections 2.1 - 4.2.

The Action-cited section 2.2 describes analysis of the radio-images of the hybridization spots. Here, Hennig suggests using duplicates spots to allow quality checks, wherein duplicate signals can be correlated, and poorly correlated or poorly reproduced signals can be discarded.

Following the clustering step (section 2.3) representative clones are selected for sequencing, and the resultant raw sequence is "cleaned" as described in the cited section 2.5.1. Hennig teaches that raw trace data pass through filtering steps, which as input takes a set of ABI trace files, and on output generates a cleaned sequence set. ABI trace files are viewable as polychromatograms depicting gated fluorescent signal intensity reads for each nucleotide base across a sequencing gel/filter. Hennig's cleaning step converts a set of such trace files into a theoretically cleaned nucleotide sequence.

So Hennig discloses an oligo-fingerprint strategy for large-scale gene expression analysis. Again we ask, how does the practitioner of Wolffe find applicable relevance in

Hennig, and to what end? Wolffe is characterizing novel regulatory sequences by using alignment tools to compare them with known sequences. Hennig is characterizing large EST libraries based on oligo fingerprinting, so as to reduce the number of clones that need to be sequenced. The Action proposes that Hennig's teachings would have allowed Wolffe to clean, remove duplicates, and perform quality checks to the raw sequence in preparation for the sequence comparison analysis. Action, p.4, lines 6-8. Clean what? Remove duplicates of what? Perform quality checks on what raw sequence? The rejection does not survive even superficial scrutiny.

Wolffe compares his identified sequences with reference sequences such as in Genbank to generate "hits", such as by using the BLAST algorithm. Of course, to the extent a Wolffe practitioner is generating original sequence, she may well seek to improve the relevance of her sequencing by sequencing multiple sample copies, and using algorithms to identify and discount artifactual sequences. This is not really analogous to what Hennig is doing: spotting duplicate probes to insure accuracy of each probe-EST correlation, but it could be argued to be general motivation to repeat data points and improve accuracy. But as much as coopting Hennig's data cleaning, removing duplicates, and performing quality checks may improve accuracy, it has not driven Wolffe's practitioner into a different line of work. suited datasets for use in our claimed methods.

We appreciate that the claimed subject matter is arcane and not easy to examine; however, we believe that the presently cited art does not provide a remotely colorable suggestion of the subject claims. We believe that our Specification provides a detailed description, analysis and distinction of prior work that those skilled in the art would find most relevant to our invention. We have laid out the features of such prior work, including the computational tools known as DRAGON, POMPOUS, Rep-X, etc., identified their deficiencies, and explained how our invention improves upon them. The nonobviousness of our invention has endured the tests of time and continued peer-review: our invention was developed and published several year ago, we know of no more-relevant prior art, and a commercial embodiment of our invention, ARROGANT, enjoys critical acclaim in this narrowly technical, but important field.

Though the cited art does not support any prima facie case under 35USC103, for good measure, we provided the Examiner with affirmative evidence in the form of an expert Declaration by Professor Garner averring to the foregoing. The Action does not controvert this evidence of record. Accordingly, the uncontroverted evidence of record demonstrates that the

cited art does not suggest the invention as claimed.

Claim 8 further requires that the recited database is one of a plurality of databases correlating the sequence identifiers of the first subset with syngeneic biopolymers, and the redundancy reducing function compares the first subset with the databases and outputs the second subset of the dataset. This claim further requires the redundancy reducing function which compares the recited first subset with multiple data sets correlating the sequence identifiers of the first subset with syngeneic biopolymers, further isolating this invention from the cited art which does not suggest or relate to a method having any such redundancy reducing function.

Claim 9 further requires that the recited parameter is either source, species, author and pathway. This claim further requires a specific type of parameter in the recited selection function, further isolating this invention from the cited art which does not suggest a method having any such selection function.

Claim 10 further requires that the recited parameter is one of a plurality of user-defined selection parameters, and the selection function applies to the second subset the parameters and outputs the third subset restricted relative to the second subset by the parameters. This claim further requires multiple selection parameters in the recited selection function, further isolating this invention from the cited art which does not suggest a method having any such selection function.

Claim 11 further requires that the recited redundancy reducing function outputs a second subset of the dataset which eliminates unique, natural complex biopolymer redundancy relative to the first subset. This claim further requires the redundancy function eliminate unique, natural complex biopolymer redundancy , further isolating this invention from the cited art which does not suggest or relate to a method having any such redundancy reducing function.

Claim 12 further requires an expansion function which searches a second database for synonyms of the sequence identifiers of the first, second or third subset. The cited art does not disclose or suggest a method comprising any such expansion function, further isolating this invention from the cited art.

Claim 14 is restricted to a computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the system comprising (a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with

syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and (b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset. The cited art does not disclose or suggest a method comprising any such merge and redundancy reducing function nor such a tabulation function, further isolating this invention from the cited art.

Claim 15 further requires that the recited merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter. This claim further restricts the merge and redundancy reducing function to a species comprising a particularly described selection function. The cited art does not disclose or suggest a method comprising any such merge and redundancy reducing function, further isolating this invention from the cited art.

Claim 16 further requires that the recited merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter, wherein the parameter is either source, author and pathway. Relative to claim 15, this claim further requires a specific type of parameter in the recited selection function, further isolating this invention from the cited art which does not suggest a method having any such merge and redundancy reducing function.

Claim 17 is restricted to a computer-based method for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the method comprising computer-implemented steps of: (a) comparing the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creating a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and (b) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset. The cited art does not disclose or suggest a method comprising any such comparing and creating and outputting steps, further isolating this invention from the cited art.

Claim 18 is restricted to a computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the system comprising: (a) an integration function which merges the dataset with a database comprising second annotations

attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and (b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations. The cited art does not disclose or suggest a method comprising any such integration and tabulation steps, further isolating this invention from the cited art.

Claim 19 further requires that the recited second annotations comprise data attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset, said data selected from the group consisting of: gene expression data, sequencing data, genotype data, polymorphism data and clinical data. Relative to claim 18, this claim further requires a specific type of data be used as second annotations in the integration step, further isolating this invention from the cited art which does not suggest a method having any such integration step.

Claim 20 is restricted to a computer-based method for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the method comprising computer-implemented steps of: (a) merging the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and linking the second annotations to the corresponding sequence identifiers of the subset; and (b) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations. The cited art does not disclose or suggest a method comprising any such merging and creating and outputting steps, further isolating this invention from the cited art.

Claim 21 is directed to the method of claim 1, discussed above, but further requires a second computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the second system comprising (a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and (b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising,

configurable by and sortable by the sequence identifiers of the subset. This claim requires that the steps of claim 1 be practiced in conjunction with a computer-based second system comprising an additional merge and redundancy reducing function and an additional a tabulation function. The cited art does not disclose or suggest a method comprising any such merge and redundancy reducing function nor such additional a tabulation function, further isolating this invention from the cited art.

Claim 22 is directed to the method of claim 1, discussed above, but further requires a second computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the second system comprising: (a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and (b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations. This claim requires that the steps of claim 1 be practiced in conjunction with a computer-based second system comprising an additional integration function and an additional a tabulation function. The cited art does not disclose or suggest a method comprising any such integration function nor such additional a tabulation function, further isolating this invention from the cited art.

Claim 23 is directed to the method of claim 1, discussed above, but further requires a second computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the second system comprising: (a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and (b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset; and, a third computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the third system comprising: (c) an integration function which merges the



dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and (d) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations. This claim requires that the steps of claim 1 be practiced in conjunction with a second computer-based system comprising an additional merge and redundancy reducing function and an additional a tabulation function and a third computer-based system comprising an additional integration function and an additional a tabulation function. The cited art does not disclose or suggest a method comprising any such second and third systems, further isolating this invention from the cited art.

Claim 24 is directed to the method of claim 1, discussed above, but further specifies that the system is none other than Applicant' proprietary ARROGANT system, as described in the Specification from p.16, line 29 - p.38, line 28. This system specifies particular and detailed mode features and input page, retrieved and output display fields. This unique, detailed combination of features is nowhere suggested in the cited art, further isolating this invention from the cited art.

## II. THE EXAMINER HAS NOT PROPERLY REJECTED CLAIMS 2 AND 4 UNDER 35USC103(a).

For dependent claims 2 and 4, the Action supplements Wolffe and Hennig with Lincoln et al. (US Pat No. 6,303,297). This additional reference does not add relevant content to the already cited art; in fact, for the cited teachings, they are largely redundant with functionalities present in well-known computational tools and databases, such as PRIMO, BLAST, and RepX. In particular, Lincoln describes a relational database for storing genetic information, and is cited for well-known uses of specific search criteria of keywords and concepts.

Though the cited art does not support any prima facie case under 35USC103, for good measure, we provided the Examiner with affirmative evidence in the form of an expert Declaration by Professor Garner averring to the foregoing. The Action does not controvert this evidence of record. Accordingly, the uncontroverted evidence of record demonstrates that the cited art does not suggest the invention as claimed.

Claim 4 further requires that the recited criterion is one of a plurality of user-defined

criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria, wherein the criteria include multiple keywords. The cited art does not disclose or suggest a method comprising any such search function, further isolating this invention from the cited art.

III. THE EXAMINER HAS NOT PROPERLY REJECTED CLAIM 5 UNDER 35USC103(a).

For dependent claim 5, the Action supplements Wolffe and Hennig with the previously cited Chin et al. (US Pat No. 6,470,277). This additional reference does not add relevant content to the already cited art; in fact, for the cited teachings, they are largely redundant with functionalities present in well-known computational tools and databases, such as PRIMO, BLAST, and RepX. In particular, Claim 5 requires that the recited dataset is GenBank, Medline or KEGG. As noted in the cited Chin et al., these are well-known sequence databases, and which provide particularly suited datasets for use in our claimed methods.

Though the cited art does not support any prima facie case under 35USC103, for good measure, we provided the Examiner with affirmative evidence in the form of an expert Declaration by Professor Garner averring to the foregoing. The Action does not controvert this evidence of record. Accordingly, the uncontroverted evidence of record demonstrates that the cited art does not suggest the invention as claimed.

IV. THE EXAMINER HAS NOT PROPERLY REJECTED CLAIM 7 UNDER 35USC103(a).

For dependent claim 7, the Action supplements Wolffe and Hennig with the previously cited MacLeod et al. (US Pat No. 6,221,600). This additional reference does not add relevant content to the already cited art; in fact, for the cited teachings, they are largely redundant with functionalities present in well-known computational tools and databases, such as PRIMO, BLAST, and RepX. In particular, Lincoln describes a relational database for storing genetic information, and is cited for well-known uses of specific search criteria of keywords and concepts. Claim 7 requires that the recited database is UniGene or LocusLink. As noted in the cited MacLeod et al., these are well-known sequence databases, and which provide particularly suited datasets for use in our claimed methods.

Though the cited art does not support any prima facie case under 35USC103, for good

measure, we provided the Examiner with affirmative evidence in the form of an expert Declaration by Professor Garner averring to the foregoing. The Action does not controvert this evidence of record. Accordingly, the uncontroverted evidence of record demonstrates that the cited art does not suggest the invention as claimed.

Appellants respectfully request reversal of the pending Final Action by the Board of Appeals.

We petition for and authorize charging our Deposit Account No.19-0750 all necessary extensions of time. The Commissioner is authorized to charge any fees or credit any overcharges relating to this communication to our Dep. Acct. No.19-0750 (order UTSD:0668).

Respectfully submitted,  
SCIENCE & TECHNOLOGY LAW GROUP



Richard Aron Osman, J.D., Ph.D., Reg.No. 36,627  
Tel(949) 218-1757; Fax(949) 218-1767

## CLAIMS ON APPEAL

1. A computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding annotations, the system comprising:
  - a) a search function which searches the annotations of the dataset according to a user-defined criterion and outputs a first subset of the dataset restricted by the criterion;
  - b) a redundancy reducing function which compares the first subset with a first database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputs a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;
  - c) a selection function which applies to the second subset a user-defined selection parameter and outputs a third subset restricted relative to the second subset by the parameter; and
  - d) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset.
2. A system according to claim 1, wherein the criterion is selected from the group consisting of a keyword and a concept.
3. A system according to claim 1, wherein the criterion is one of a plurality of user-defined criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria.
4. A system according to claim 1, wherein the criterion is one of a plurality of user-defined criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria, wherein the criteria include multiple keywords.
5. A system according to claim 1, wherein the dataset is selected from the group consisting of GenBank, Medline and KEGG.

6. A system according to claim 1, wherein the dataset is one of a plurality of datasets, and the search function searches the annotations of the datasets according to the user-defined criterion and outputs a first subset of the datasets restricted by the criterion.
7. A system according to claim 1, wherein the database is selected from the group consisting of UniGene and LocusLink.
8. A system according to claim 1, wherein the database is one of a plurality of databases correlating the sequence identifiers of the first subset with syngeneic biopolymers, and the redundancy reducing function compares the first subset with the databases and outputs the second subset of the dataset.
9. A system according to claim 1, wherein the parameter is selected from the group consisting of source, species, author and pathway.
10. A system according to claim 1, wherein the parameter is one of a plurality of user-defined selection parameters, and the selection function applies to the second subset the parameters and outputs the third subset restricted relative to the second subset by the parameters.
11. A system according to claim 1, wherein the redundancy reducing function outputs a second subset of the dataset which eliminates unique, natural complex biopolymer redundancy relative to the first subset.
12. A system according to claim 1, further comprising an expansion function which searches a second database for synonyms of the sequence identifiers of the first, second or third subset.
13. A computer-based method for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding annotations, the method comprising computer-implemented steps of:

- a) searching with a computer the annotations of the dataset according to a user-defined criterion and outputting a first subset of the dataset restricted by the criterion;
- b) comparing with the computer the first subset with a database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputting a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;
- c) applying to the second subset a user-defined selection parameter and outputting a third subset restricted relative to the second subset by the parameter; and
- d) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset

14. A computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the system comprising:

- a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and
- b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset.

15. A system according to claim 14, wherein the merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter.

16. A system according to claim 14, wherein the merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter, wherein the parameter is selected from the group consisting of source, author and pathway.

17. A computer-based method for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the method comprising computer-implemented steps of:

a) comparing the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creating a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset.

18. A computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

19. A system according to claim 18, wherein the second annotations comprise data attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset, said data selected from the group consisting of: gene expression data, sequencing data, genotype data, polymorphism data and clinical data.

20. A computer-based method for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the method comprising computer-implemented steps of:

a) merging the dataset with a database comprising second annotations attributable to and

correlated with at least a subset of the sequence identifiers or sequences of the dataset and linking the second annotations to the corresponding sequence identifiers of the subset; and

b) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

21. A system according to claim 1, further comprising:

a second computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the second system comprising:

a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset.

22. A system according to claim 1, further comprising:

a second computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the second system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.



23. A system according to claim 1, further comprising:

a second computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the second system comprising:

a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset; and,

a third computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the third system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

24. A system according to claim 1, wherein the system is ARROGANT.